South Orange-Maplewood Academic Performance on the 2005 School Report Card: An Analysis

Wayne Eastman
430 Meeker Street
South Orange, NJ 07079
weastman@andromeda.rutgers.edu

Contents--

I. Summary

II. SO-M Performance
   a. Columbia, middle schools, and elementary schools
   b. Summary of performance and change in SO-M elementary schools
   c. Relative to certain schools in Abbott districts
   d. Relative to Livingston, Millburn, Montclair, and West Orange High Schools

III. Disaggregated SO-M Performance

IV. Practical Judgments and Value Judgments in the Use of Aggregated and Disaggregated Data

V. Notes on Methods and Assumptions

VI. Tables (separate document)
I. Summary

What follows is an analysis of the performance of SO-M schools on the recently released 2005 New Jersey Report Card. The report focuses on performance relative to other schools in our district factor group on the state-mandated ASK3, ASK4, GEPA, and HSPA. It also reports data for public schools in Newark, East Orange, Orange, and Irvington that have achieved better or equal results on one or more of the state tests in comparison to one or more SO-M schools. For Columbia High School, the report also considers the SATs and the APs, and includes a comparison of Columbia results to those in Livingston, Millburn, Montclair, and West Orange.

The major findings of the analysis can be summarized as follows:

1. Performance on the HSPA and GEPA relative to our DFG is below average in SO-M (with the exception of a performance on GEPA science at SOMS that is much below average). (See Section V for an explanation of average, average -, below average, much below average, and all the other performance ratings used in this report.)

2. For performance on the ASK3 and ASK4 relative to our DFG, South Mountain is a bright spot. ASK performance ranges from below average to very much below average at Clinton; from average – to much below average at Jefferson; from much below average to very much below average at Seth Boyden; from average to above average at South Mountain; and from average – to much below average at Tuscan.

3. Change in SO-M from 2004 to 2005 on the state tests relative to the state average change varied. Columbia was moderately above average in change on the HSPA; MMS was below average and SOMS was above average on the GEPA. Among the elementary schools, Clinton and South Mountain were better than the state average, Jefferson and Seth Boyden were worse, and Tuscan was average.

4. There are 41 public schools (including 5 charter schools) in Newark, East Orange, Orange, or Irvington with results on one or more of the state tests that are better than those in one or more SO-M schools. 20 elementary schools in the four Abbott districts outperformed Seth Boyden on a majority of the ASK tests and 1 was comparable; 14 schools outperformed Clinton, with 1 comparable, 8 outperformed Jefferson, with 3 comparable; 5 schools outperformed Tuscan, with 2 comparable; 1 school outperformed South Mountain.

5. SAT results at Columbia are considerably below our factor group, and also trail Montclair High School.
6. There is a disparity between HSPA performance at Columbia and SAT performance. On the much more important SAT, performance is considerably worse than on the HSPA.

7. SAT change from 2003-2005 at Columbia High School lagged the state and also lagged Livingston, Millburn, and Montclair, while coming slightly ahead of West Orange.

8. AP participation and 3+ scoring levels at Columbia significantly lag those in Montclair, Livingston, and Millburn, and are comparable to those in West Orange.

9. Reliance on the non-rigorous SRA alternative to HSPA graduation is higher at Columbia than the state average, much higher than at Montclair, and slightly lower than in West Orange.

10. The disaggregated results show a level of performance by black students that moderately lags black students in the DFG. Black families in our towns have a high level of affluence relative to black students elsewhere in the state (and to a lesser degree in the DFG—the other district in the DFG with a large number of black students in the DFG is Montclair, in which black affluence is significantly less than in SO-M). The disaggregated results thus indicate a disparity between the relative level of performance by SO-M black students that one would expect based on socio-economic status and lower actual performance.

11. Disaggregated results on state tests show a level of performance by white students that is above average for the DFG, with stronger performance at the high school and middle school level than at the elementary school level. However, performance by whites is not uniformly good, with average or below average performance by white students relative to the DFG on some state tests at some schools (a considerable amount of missing data for elementary schools makes it difficult to draw strong conclusions here and with respect to other matters involving disaggregated data). Finally, the comparative weakness at Columbia on the important SAT relative to the HSPA appears to affect whites as well as blacks.

12. On state tests, the size of the achievement gap relative to the DFG average is much larger (language arts) and larger (math) at Columbia High School; much larger (LA) and very much larger (M) at MMS; average (LA) and slightly below average (M) at SOMS. At the elementary schools, the size of the gap relative to the DFG is about average at South Mountain and Clinton, and substantially larger than average at Jefferson and Seth Boyden; Tuscan data are currently unavailable.
II.a. Academic Performance of SO-M Schools

1. Columbia High School

HSPA L: Below average
HSPA M: Below average
HSPA L 2004-2005 change: Average +
HSPA M 2004-2005 change: Above average
SRA (non-HSPA) graduation: CHS 4 points worse
SAT M avg.: CHS 45 points worse
SAT M 25th percentile: CHS 60 points worse
SAT M 50th percentile: CHS 50 points worse
SAT M 75th percentile: CHS 30 points worse
SAT V avg.: CHS 39 points worse
SAT V 25th percentile: CHS 70 points worse
SAT V 50th percentile: CHS 40 points worse
SAT V 75th percentile: CHS 30 points worse
SAT M avg. change 2003-2005: CHS 16 points worse
SAT M 25th percentile change 2003-2005: CHS 2 points worse
SAT M 50th percentile change 2003-2005: CHS 22 points worse
SAT M 75th percentile change 2003-2005: CHS 21 points worse
SAT V avg. change 2003-2005: CHS 10 points worse
SAT V 25th percentile change 2003-2005: CHS 7 points worse
SAT V 50th percentile change 2003-2005: CHS 11 points worse
SAT V 75th percentile change 2003-2005: CHS 14 points worse
AP % vs. state: CHS 6 points better
% of school with 3+ AP: 17
HSPA average--CHS leads the state by over one quarter of a standard deviation in LA
and about .6 of a standard deviation on math.
SAT average—CHS leads the state by about one tenth of a standard deviation on both V
and M

 Comments: Columbia’s HSPA performance is below average; year-to-year HSPA change
relative to the state is moderately above average. Columbia’s SAT performance is weak,
with weaker performance at lower percentiles than at higher percentiles. SAT
performance at Columbia is worse than HSPA performance, as indicated by the fact that
Columbia tests only slightly above the state average on the SAT verbal and math while
testing somewhat above average on the HSPA language arts and considerably above
average on the HSPA math. SAT 2003-2005 change at Columbia is below average
relative to the state, especially at higher percentiles. The higher percentage of non-
rigorous SRA graduation at Columbia than in N.J. as a whole (18.2% compared to a state
average of 14.2%) is a basis for concern. Columbia has a significantly higher AP
participation rate than the state average, but this is hard to interpret given the high factor
group of SO-M.

2. Maplewood Middle School
GEPA L: Below average  
GEPA M: Below average  
GEPA S: Below average  
GEPA L 2004-2005 change: Below average  
GEPA M 2004-2005 change: Below average  
GEPA S 2004-2005 change: Below average  

Comments: Maplewood shows fairly weak performance. All three categories in SOMS and MMS are below average, but SOMS is slightly better than MMS in language arts and math and MMS is slightly better in science. MMS has a consistent pattern of negative year-to-year change, in contrast to the consistent positive SOMS year-to-year pattern.

(Although this section is devoted to aggregate data, it is worth noting that SOMS has GEPA L figures of 99 whites and 154 blacks, while MMS has GEPA L figures of 98 whites and 121 blacks. Given the slightly better performance of SOMS, this counters the assumption that superior performance will be associated with a whiter school.)

3. South Orange Middle School

GEPA L: Below average  
GEPA M: Below average  
GEPA S: Much below average  
GEPA L 2004-2005 change: Above average  
GEPA M 2004-2005 change: Above average  
GEPA S 2004-2005 change: Above average  

Comments: SOMS shows fairly weak performance but a consistent year-to-year improvement relative to the state.

4. Clinton Elementary School

Test Performance:  
ASK3 L: Below average  
ASK3 M: Very much below average  
ASK4 L: Below average  
ASK4 M: Much below average  
ASK4 S: Much below average  
ASK3 L 2004-2005 change: Much above average  
ASK3 M 2004-2005 change: Below average  
ASK3 M 2004-2005 change: Average +  
ASK4 L 2004-2005 change: Average +  
ASK4 M 2004-2005 change: Much above average  
Number economically disadvantaged 2005 ASK4: 14  
Number economically disadvantaged 2004 ASK4: 14
Comments: Clinton performance is not good. (Although Clinton is less affluent than certain other areas of SO-M and the school has a considerably larger number of students in families where English is not the first language, the number of economically disadvantaged students per the state report card is modest.) On year-to-year change, the pattern is mixed but generally positive, with strong gains in grade 3 LA and grade 4 math.

5. Jefferson Elementary School

Test Performance:
ASK3 L: Below average
ASK3 M: Much below average
ASK4 L: Below average
ASK4 M: Much below average
ASK4 S: Average -
ASK3 L 2004-2005 change: Average
ASK3 M 2004-2005 change: Below average
ASK4 L 2004-2005 change: Below average
ASK4 M 2004-2005 change: Much below average
Number economically disadvantaged 2005 ASK4: less than 10
Number economically disadvantaged 2004 ASK4: 16

Comments: Especially given the prevailing high levels of affluence of families in the Marshall-Jefferson pairing, performance on the ASK3 and ASK4 is weak, even allowing for the comparatively high number of students with disabilities. It is also a concern that change from 2003-2004 to 2004-2005 lags behind state change. The comparatively large number of disabled students cannot be used to explain Jefferson’s below average performance in regard to change. Also note that there are no results from Marshall because the school is K-2 and statewide testing starts in third grade.

6. Seth Boyden Demonstration School

Test Performance:
ASK3 L: Much below average
ASK3 M: Very much below average
ASK4 L: Much below average
ASK4 M: Very much below average
ASK4 S: Very much below average
ASK3 L 2004-2005 change: Below average
ASK3 M 2004-2005 change: Average +
ASK4 L 2004-2005 change: Below average
ASK4 M 2004-2005 change: Very much below average
Number economically disadvantaged 2005 ASK4: 18
Number economically disadvantaged 2004 ASK4: 12
Comments: Seth Boyden performance is weak. (Although Seth Boyden’s neighborhood is the least affluent part of SO-M, the number of economically disadvantaged students per the state report card is not high.) On year-to-year change, the pattern is also not good, with a very poor change result in fourth grade math standing out.

7. South Mountain Elementary School

Performance:
ASK3 L: Average
ASK3 M: Average +
ASK4 L: Average
ASK4 M: Above average
ASK4 S: Average +
ASK3 L 2004-2005 change: Average
ASK3 M 2004-2005 change: Above average
ASK4 L 2004-2005 change: Average
ASK4 M 2004-2005 change: Much above average
Number economically disadvantaged 2005: NA
Number economically disadvantaged 2004: less than 10

Comments: Performance on the ASK3 and ASK4 is reasonably good. Change from 2003-2004 to 2004-2005 is better than the state. South Mountain performance is a bright spot in a district not characterized by good performance. Curricular and teaching practices employed at South Mountain should be considered as a potential model for other elementary schools in the district.

(Although this section of the report is devoted to aggregate data, it is worth noting the racial composition of South Mountain, which was 47 whites and 44 blacks for ASK4. For Jefferson, the comparable numbers were 66 whites and 57 blacks; for Seth Boyden 28 whites and 58 blacks; for Clinton ASK3 23 whites and 44 blacks; and for Tuscan 51 whites and 29 blacks (2004 data for Tuscan). Accordingly, the South Mountain numbers cannot readily be explained based on the idea that the school differs sharply in its racial composition from other district elementary schools. It should also be noted that the South Mountain catchment area is not limited to high income areas; it extends down from the school in a wedge to include parts of downtown South Orange and eastern South Orange near Seton Hall that are moderate income by local standards.)

8. Tuscan Elementary School

Test Performance:
ASK3 L: Average -
ASK3 M: Average -
ASK4 L: Much below average
ASK4 M: Much below average
ASK4 S: Average
ASK3 L 2004-2005 change: Above average
ASK3 M 2004-2005 change: Much above average
ASK4 L 2004-2005 change: Below average
ASK4 M 2004-2005 change: Much below average
Number economically disadvantaged 2005 ASK4: 11
Number economically disadvantaged 2004 ASK4: 15

Comments: Tuscan shows a very mixed pattern. Performance is acceptable though not strong among third graders, but is weak among fourth graders. A similar pattern applies to change, in which there are very favorable results among third graders and very unfavorable ones among fourth graders. Although the pattern may be partially the result of the 2004-2005 Tuscan fourth grade class being weaker than its predecessor or successor classes, there is reason for concern and for scrutiny of the fourth grade situation at Tuscan.
II. b. Summary of test performance (relative to DFG) and change (relative to state) in SO-M elementary schools

<table>
<thead>
<tr>
<th>School</th>
<th>Performance</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinton</td>
<td>-9</td>
<td>+3.5</td>
</tr>
<tr>
<td>Jefferson</td>
<td>-8</td>
<td>-5.0</td>
</tr>
<tr>
<td>Seth Boyden</td>
<td>-13</td>
<td>-4.5</td>
</tr>
<tr>
<td>South Mountain</td>
<td>+2</td>
<td>+3.0</td>
</tr>
<tr>
<td>Tuscan</td>
<td>-5.5</td>
<td>0</td>
</tr>
</tbody>
</table>

Explanation: Scores = sum of performances on ASK3 and ASK4. 
Average = 0, Average + = +.5, Average - = -.5, Above average = +1, Below average = -1, Much above average = +2, Much below average = -2, Very much above average = +3, Very much below average = -3.
II.c. SO-M Performance Relative to Certain Schools in Nearby Abbott Districts

Overall, as would be expected, SO-M schools substantially outperform schools in Abbott districts. This section focuses on certain schools in four Abbott districts near us—Newark, East Orange, Orange, and Irvington—that have relatively good levels of student performance on at least some tests; the schools to the best of the author’s knowledge have nonselective admissions policies unless it is noted that admission is selective. The aim of this section is to make readers aware that students in these schools have in some cases performed on tests at levels exceeding (or in a few cases equaling) those in SO-M. The policy implication is that it is highly questionable for SO-M schools, all of which have much higher student socio-economic statuses than those of the Abbott district schools, to use the composition of their student bodies as a basis to rationalize sub-par test performance. (It should be noted that the measure of performance used in this study takes both advanced and partially proficient performance into account, which in some cases favors SO-M schools over the Abbott district schools. For example, Fifteenth Avenue School in Newark has a lower % of partially proficient students on the ASK3 M than Jefferson and Seth Boyden, but is not shown as doing better than the two SO-M schools because of the advantage Jefferson and Seth Boyden have over Fifteenth Avenue in advanced proficient %.)

1. Abington Ave.—Newark.

The ASK3 math and language arts performance of Abington Avenue students is better than that in all SO-M elementary schools except for South Mountain. The ASK4 Abington Ave. language arts and math performance is also better than all SO-M elementary schools except South Mountain.

2. Alexander Ave.—Newark

ASK3 M results are better at Alexander Ave. than at Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than at Seth Boyden and very similar to those at Clinton and Jefferson.

3. Ann St.—Newark

ASK3 LA results are better than at Clinton, Jefferson, and Seth Boyden; ASK3 M results are better than at all SO-M elementary schools except South Mountain. ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan, and very similar to Clinton’s; ASK4 M results are better than all SO-M schools except South Mountain’s; ASK4 S results are better than Clinton and Seth Boyden.

4. Branch Brook—Newark

ASK3 M results are better than Clinton, Jefferson, and Seth Boyden.
5. Fifteenth Avenue—Newark

ASK4 M results are better than Seth Boyden.

6. First Avenue—Newark

ASK4 LA and math results are better than all SO-M schools except South Mountain.

7. Fourteenth Ave.—Newark

ASK3 M, ASK4 M, and ASK4 S results are better than all SO-M schools.

8. Harriet Tubman—Newark

ASK3 results are better than Seth Boyden and equal to Jefferson; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden, and very similar to Tuscan. ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden.

9. Lafayette St.—Newark

ASK3 LA results are better than Seth Boyden’s; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden; ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan; ASK4 M results are better than Clinton, Seth Boyden, and Tuscan; ASK4 S results are better than Clinton and Seth Boyden.

10. McKinley—Newark

ASK4 LA results are better than all SO-M schools but South Mountain; ASK4 M results are better than Clinton, Seth Boyden, and Tuscan; ASK4 S results are than Clinton and Seth Boyden.

11. Miller St.—Newark

ASK4 M results are better than Seth Boyden.

12. Mount Vernon—Newark

ASK3 LA results are better than Seth Boyden; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than all SO-M schools except South Mountain; ASK4 M results are better than Seth Boyden.
13. Newton St.--Newark

ASK4 LA results are comparable to Seth Boyden; ASK4 M results are better than Seth Boyden; ASK4 S results are better than Clinton and Seth Boyden.

14. Oliver St.—Newark

ASK3 LA results are better than Seth Boyden and comparable to Jefferson; ASK3 M results are better than all SO-M schools except South Mountain. ASK4 LA results are better than all SO-M schools except South Mountain and Clinton; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden and comparable to Jefferson.

15. Ridge St.—Newark

ASK4 M results are better than Seth Boyden.

16. Roseville Ave.—Newark

ASK3 M results are better than all SO-M schools except South Mountain and Tuscan; ASK4 M results are better than all SO-M schools except South Mountain.

17. Science High—Newark (selective admission)

HSPA LA and M results are better than Columbia.

18. Speedway Ave.—Newark

ASK4 M results are better than all SO-M schools except Jefferson and South Mountain.

19. Sussex Ave.—Newark

ASK3 M results are better than all SO-M schools except South Mountain and Tuscan;

20. Warren St.—Newark

ASK4 M results are better than Seth Boyden.

21. Wilson Ave.—Newark

ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 M results are better than Clinton, Seth Boyden, and Tuscan

22. Dionne Warwick Ave.—East Orange
ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden.

23. Dr John Howard Jr U-S-E—East Orange

ASK4 LA results are comparable to Seth Boyden; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden.

24. Fourth Avenue—East Orange

ASK3 M results are better than Clinton, Jefferson, and Seth Boyden; ASK4 M results are comparable to Seth Boyden.

25. Gordon Parks Academy—East Orange

ASK3 results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are comparable to Seth Boyden; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden.

26. J. Garfield Jackson Sr. Academy

ASK3 LA results are better than Seth Boyden; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than Seth Boyden and comparable to Tuscan; ASK4 M results are better than Clinton, Seth Boyden, and Tuscan.

27. Johnnie Cochran Academy

ASK3 LA results are better than Jefferson and Seth Boyden and comparable to Clinton; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden, and comparable to Tuscan. ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan, and comparable to Clinton; ASK M results are better than all SO-M schools except South Mountain.

28. Langston Hughes Academy

ASK3 M results are better than Clinton, Jefferson, and Seth Boyden; ASK4 math results are better than Seth Boyden and comparable to Clinton.

29. Washington Academy—East Orange

ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA, M, and S results are better than all SO-M schools except South Mountain.
30. Whitney E. Houston Academy—East Orange

ASK4 LA results are better than Seth Boyden and comparable to Tuscan; ASK4 M results are better than Seth Boyden.

31. Central School—Orange

ASK3 LA results are better than Seth Boyden; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than Seth Boyden and Tuscan and comparable to Jefferson; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Clinton and Seth Boyden.

32. Cleveland Street—Orange

ASK3 M results are better than all SO-M schools; ASK4 M results are better than Seth Boyden and comparable to Clinton.

33 Heywood—Orange

ASK3 LA results are better than Seth Boyden.

34 Oakwood Ave.—Orange

ASK3 M results are better than all SO-M schools. ASK4 LA results are comparable to Seth Boyden; ASK4 M results are better than all SO-M schools except South Mountain; ASK4 S results are better than Seth Boyden and comparable to Clinton.

35. Park Ave.—Orange

ASK3 LA results are better than Jefferson and Seth Boyden; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than Jefferson, Seth Boyden, and Tuscan, and comparable to Clinton; ASK4 M results are better than all SO-M schools except South Mountain.

36. Chancellor Ave.—Irvington

ASK3 LA results are comparable to Seth Boyden; ASK4 LA results are better than all district schools except for South Mountain. ASK4 M results are better than Clinton, Seth Boyden, and Tuscan, and are comparable to Jefferson; ASK4 S results are better than Clinton and Seth Boyden.

37. East Orange Community Charter School—East Orange

ASK3 LA results are better than Seth Boyden and comparable to Jefferson; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better
than all SO-M schools except South Mountain; ASK4 M results are better than Clinton, Seth Boyden, and Tuscan; ASK4 S results are better than Clinton and Seth Boyden.

38. Gray Charter School—Newark

ASK3 LA results are better than all SO-M schools except South Mountain and ASK3 M results are better than all SO-M schools. ASK4 LA results are better than all SO-M schools except South Mountain; ASK4 M results are comparable to South Mountain and better than all other SO-M schools; ASK4 S results are better than Clinton and Seth Boyden and comparable to Jefferson.

39. Lady Liberty Charter School—Newark

ASK3 LA results are better than Seth Boyden.

40. Marion Thomas Charter School—Newark

ASK3 LA results are better than Jefferson and Seth Boyden; ASK3 M results are better than Clinton, Jefferson, and Seth Boyden. ASK4 LA results are better than Seth Boyden and comparable to Jefferson and Tuscan.

41. Robert Treat Charter School—Newark

ASK3 LA and M results are better than all SO-M schools except South Mountain. ASK4 LA results are better than all SO-M schools; ASK4 M results are comparable to South Mountain and better than all other SO-M schools; ASK S results are better than Clinton, Jefferson, and Seth Boyden, and comparable to Tuscan. GEPA LA results are better than MMS and SOMS; GEPA M results are comparable to MMS.
II.d. SO-M Performance Relative to Livingston, Millburn, Montclair, and West Orange High Schools

1. Columbia High School relative to Livingston High School

HSPA L: CHS 21 points worse
HSPA M: CHS 28 points worse
HSPA L 2004-2005 change: CHS 11 points better
HSPA M 2004-2005 change: CHS 14 points better
SRA (non-HSPA) graduation: CHS 15 points worse
SAT M avg.: CHS 83 points worse
SAT M 25th percentile: CHS 80 points worse
SAT M 50th percentile: CHS 110 points worse
SAT M 75th percentile: CHS 80 points worse
SAT V avg.: CHS 63 points worse
SAT V 25th percentile: CHS 30 points worse
SAT V 50th percentile: CHS 50 points worse
SAT V 75th percentile: CHS 40 points worse
SAT M avg. change 2003-2005: CHS 22 points worse
SAT M 25th percentile change 2003-2005: Equal
SAT M 50th percentile change 2003-2005: CHS 30 points worse
SAT M 75th percentile change 2003-2005: CHS 30 points worse
SAT V avg. change 2003-2005: CHS 23 points worse
SAT V 25th percentile change 2003-2005: CHS 30 points worse
SAT V 50th percentile change 2003-2005: CHS 10 points worse
SAT V 75th percentile change 2003-2005: CHS 30 points worse
% taking SAT: CHS 10 points worse
AP %: CHS 9 points worse
% of school with 3+ AP: CHS 12 points worse

Comments: The comparison favors Livingston over Columbia across the board, with the exception of HSPA year-to-year improvement. One might try to explain this result on the basis of higher SES in Livingston, but the state’s factor group system places both Livingston and SO-M in factor group I. (For discussion of disaggregated data, see Section III.) On the SAT, Livingston performs much better, and also has a better change from 2003-2005, along with a higher participation rate (99% versus 89%). Livingston has a much lower rate (3% versus 18%) of less rigorous SRA graduation. Livingston’s HSPA figures are considerably better than Columbia’s, but Columbia shows more 2004-2005 improvement. (It should be noted that performance on the SAT, where Columbia lags Livingston substantially, is considerably more important for most students than performance on the HSPA, since the SAT unlike the HSPA is a major determinant in college admissions.) On both AP participation (30% LHS, 21% CHS) and ratio of AP scores of 3+ to the student population (29% LHS, 17% CHS), the comparison strongly favors Livingston.
2. Columbia High School relative to Millburn High School

HSPA L: CHS 34 points worse
HSPA M: CHS 37 points worse
HSPA L 2004-2005 change: CHS 5 points better
HSPA M 2004-2005 change: CHS 10 points better
SRA (non-HSPA) graduation: CHS 17 points worse
SAT M avg.: CHS 102 points worse
SAT M 25th percentile: CHS 130 points worse
SAT M 50th percentile: CHS 110 points worse
SAT M 75th percentile: CHS 80 points worse
SAT V avg.: CHS 91 points worse
SAT V 25th percentile: CHS 120 points worse
SAT V 50th percentile: CHS 105 points worse
SAT V 75th percentile: CHS 80 points worse
SAT M avg. change 2003-2005: CHS 18 points worse
SAT M 25th percentile change 2003-2005: CHS 30 points worse
SAT M 50th percentile change 2003-2005: CHS 10 points worse
SAT M 75th percentile change 2003-2005: Equal
SAT V avg. change 2003-2005: CHS 7 points worse
SAT V 25th percentile change 2003-2005: CHS 20 points worse
SAT V 50th percentile change 2003-2005: CHS 20 points worse
SAT V 75th percentile change 2003-2005: CHS 10 points better
% taking SAT: CHS 16 points worse
AP %: CHS 19 points worse
% of school with 3+ AP: CHS 21 points worse

Comments: The comparison favors Millburn over Columbia, with the exception of HSPA year-to-year improvement. Given higher SES in Millburn (factor group J instead of SO-M’s factor group I), better results for Millburn are in one sense to be expected, but given the proximity of Millburn to our communities and its similarities in certain respects, the comparisons are worth noting. On the SAT, Millburn performs much better than Columbia, and also has a somewhat better change from 2003-2005, though not at the 75th percentile, and also has a higher participation rate. Millburn has a much lower rate (1% versus 18%) of less rigorous SRA graduation. Millburn’s HSPA figures are much better than Columbia’s, but Columbia shows more 2004-2005 improvement. (It should be noted that performance on the SAT, where Columbia lags Millburn very substantially, is considerably more important for most students than performance on the HSPA, since the SAT unlike the HSPA is a major determinant in college admissions.) On both AP participation (40% MHS, 21% CHS) and ratio of AP scores of 3+ to the student population (38% MHS, 17% CHS), the comparison very strongly favors Millburn.

3. Columbia High School relative to Montclair High School

HSPA L: CHS 3 points worse
HSPA M: CHS 6 points better
HSPA L 2004-2005 change: CHS 3 points worse
HSPA M 2004-2005 change: CHS 8 points better
SRA (non-HSPA) graduation: CHS 10 points worse
SAT M avg.: CHS 14 points worse
SAT M 25th percentile: Equal
SAT M 50th percentile: CHS 25 points worse
SAT M 75th percentile: CHS 20 points worse
SAT V avg.: CHS 30 points worse
SAT V 25th percentile: CHS 30 points worse
SAT V 50th percentile: CHS 50 points worse
SAT V 75th percentile: CHS 40 points worse
SAT M avg. change 2003-2005: CHS 35 points worse
SAT M 25th percentile change 2003-2005: CHS 15 points worse
SAT M 50th percentile change 2003-2005: CHS 45 points worse
SAT M 75th percentile change 2003-2005: CHS 40 points
SAT V avg. change 2003-2005: CHS 26 points worse
SAT V 25th percentile change 2003-2005: CHS 10 points worse
SAT V 50th percentile change 2003-2005: CHS 40 points worse
SAT V 75th percentile change 2003-2005: CHS 35 points worse
% taking SAT: CHS 3 points better
AP %: CHS 14 points worse (21% vs. 35%)
% of school with 3+ AP: CHS 4 points worse (17% vs. 21%)

Comments: The comparison favors Montclair High School over Columbia High School quite strongly on the SAT, with Montclair both performing better, especially at higher percentiles, and having a better change from 2003-2005; the slightly higher participation rate at CHS is the only SO-M positive. The comparison favors Montclair over Columbia very strongly on HSPA graduation as opposed to less rigorous SRA graduation; Columbia has 18.2% of its students graduating via SRA, while Montclair has a much lower 8.6%. On the HSPA the picture is more mixed, with Columbia showing better relative results in math and Montclair in language arts. (It should be noted that performance on the SAT, where Columbia lags Montclair, is considerably more important than performance on the HSPA; the SAT unlike the HSPA is a major determinant in college admissions, and the SRA procedure allows students unable to cope with the HSPA to graduate in any case.) On AP participation, the comparison strongly favors Montclair (35% to 21%); on the ratio of AP scores of 3+ to the student population, the comparison is also in favor of Montclair, though by a lesser ratio (21% to 17%).

4. Columbia High School relative to West Orange High School

HSPA L: CHS 8 points better
HSPA M: CHS 23 points better
HSPA L 2004-2005 change: CHS 1 point better
HSPA M 2004-2005 change: CHS 3 points better
SRA (non-HSPA) graduation: CHS 2 points better
SAT M avg.: CHS 30 points better
SAT M 25th percentile: CHS 20 points better
SAT M 50th percentile: CHS 20 points better
SAT M 75th percentile: CHS 50 points better
SAT V avg.: CHS 36 points better
SAT V 25th percentile: CHS 10 points better
SAT V 50th percentile: CHS 50 points better
SAT V 75th percentile: CHS 65 points better
SAT M avg. change 2003-2005: CHS 6 points better
SAT M 25th percentile change 2003-2005: CHS 20 points better
SAT M 50th percentile change 2003-2005: Equal
SAT M 75th percentile change 2003-2005: CHS 30 points better
SAT V avg. change 2003-2005: CHS 7 points better
SAT V 25th percentile change 2003-2005: CHS 10 points worse
SAT V 50th percentile change 2003-2005: CHS 10 points better
SAT V 75th percentile change 2003-2005: CHS 25 points worse
% taking SAT: WOHS 1 point better
AP %: CHS 1 point worse
% of school with 3+ AP: Equal

Comments: The comparison favors Columbia High School over West Orange High School, with the exception of AP participation and results, in which the two schools are very similar. A better result for Columbia is in one sense to be expected, given that West Orange is in a lower factor group, reflecting a less affluent and highly educated population, but is nonetheless worth reporting, since West Orange is geographically close to and in a number of ways similar to SO-M. West Orange’s HSPA and SAT performance lag Columbia’s; the West Orange HSPA and SAT change rates slightly lag Columbia’s, as does its SRA graduation rate (a negative indicator). West Orange leads Columbia (by a very tiny margin) in % of students taking the SAT, is slightly ahead of Columbia in AP participation % and is equal in the ratio of 3+ AP results to school population.
III. Disaggregated SO-M performance

a. Columbia High School

**HSPA LA**
- Whites—Much better than average
- Blacks—Average –
- Size of gap: Much larger than average

**HSPA M**
- Whites—Much better than average
- Blacks—Average +
- Size of gap: Larger than average

**HSPA LA change W**—Better than average
**HSPA LA change B**—Somewhat below average

**HSPA M change W**—Average
**HSPA M change B**—Average

SAT M and SAT V—The data are not disaggregated for these important tests. To make a good guess at the subgroup performance on these tests, it would be helpful to have information not shown on the report card, such as raw scores and standard deviations of HSPA results. In the judgment of this author, it is highly likely that white as well as black Columbia HS students do not do as well on the SAT relative to the DFG and other suburban Essex schools as they do on the HSPA. If that is the case, there is less reason for complacency about the performance of whites in the district than might be suggested by the HSPA results shown above.

In addition to the results reported in Section II showing that Columbia High School students as a whole score only slightly above the state on both parts of the SAT while doing considerably better on the SAT, the comparison with Montclair is interesting. For 2004-2005, Columbia is considerably behind Montclair on SAT V and M performance but only slightly behind Montclair on the HSPA LA and ahead of Montclair on HSPA M (see Section IId for the comparison of Columbia and Montclair High School). A reasonable inference from the Montclair as well as the statewide data is that on the more important SAT, white as well as black students at Columbia are not doing as well relative to those in Montclair and elsewhere in the state as they are on the HSPA.

Comments: Achievement gaps at Columbia on the HSPA were substantially larger than the DFG average. Performance by whites was much above average; performance by blacks ranged from average + on M to average - on L. 2004-2005 change relative to the state was average for both whites and blacks. As noted in more detail above, the weaker performance of Columbia students on the SAT than the HSPA is a concern for both white and black students.
b. Maplewood Middle School

GEPA L W—Above average
GEPA L B—Below average
Size of gap—Much larger than average

GEPA M W—Very much better than average
GEPA M B—Average
Size of gap—Very much larger than average

GEPA L W change—Average
GEPA L B change—Below average

GEPA M W change—Above average
GEPA M B change—Below average

Comments: Achievement gaps at MMS on the GEPA considerably exceeded those in the DFG. Performance by whites ranged from above average on L to very much above average on M; performance by blacks ranged from average on M to below average on L. 2004-2005 change relative to the state was moderately positive for whites but below average for blacks.

c. South Orange Middle School

GEPA L W—Average +
GEPA L B—Average +
Size of gap—Average

GEPA M W—Above average
GEPA M B—Much better than average
Size of gap—Somewhat smaller than average

GEPA L W change—Average
GEPA L B change—Above average

GEPA M W change—Above average
GEPA M B change—Very much better than average

Comments: Achievement gaps relative to the DFG at SOMS on the GEPA were average on LA and somewhat less than average on M. Performance by whites ranged from somewhat above average on L to above average on M; performance by blacks ranged from somewhat above average on L to much better than average on M. 2004-2005 change relative to the state was moderately positive for whites and strongly positive for
Comparing SOMS with MMS: SOMS shows a considerably smaller achievement gap. Though one would predict a somewhat smaller gap because of higher black affluence in SO than in M, the smaller gap is SOMS is nonetheless worth noting, especially in light of gaps in the district that appear large given the relatively equal economic position of blacks and whites in our towns. Additionally, the better year to year change for blacks in SOMS than in MMS, which cannot be attributed to socioeconomic factors, is worth noting.

d. Clinton Elementary School

ASK3 LA W—Average +
ASK3 LA B—Average
Size of gap—Worse than average

ASK3 M W—Below average
ASK3 M B—Average
Size of gap—Better than average

ASK3 LA W change—Very much better than average
ASK3 LA B change—Average +

ASK3 W change—Below average
ASK3 B change—Below average

ASK4 tests: 2005 disaggregated data unavailable

e. Jefferson Elementary School

ASK3 LA W—Average
ASK3 LA B—Below average
Size of gap—Worse than average

ASK3 LA W change—Much better than average
ASK3 LA B change—Below average

ASK3 M W—Average +
ASK3 M B—Very much below average
Size of gap—Very much worse than average

ASK3 M W change—Below average
ASK3 M B change—Very much below average

ASK4 LA W—Average -
ASK4 LA B—Below average
Size of gap—Average -
ASK4 LA W change—Worse than average
ASK4 LA B change—Better than average

ASK4 M W—Much better than average
ASK4 M B—Below average
Size of gap—Very much worse than average

ASK4 M W change—Average -
ASK4 M B change—Much below average

ASK4 S W—Above average
ASK4 S B—Average -
Size of gap—Much worse than average
f. Seth Boyden Demonstration School

ASK3 L W—Unavailable
ASK3 L B—Unavailable
Size of gap—Unavailable

ASK3M W—Much better than average
ASK3M B—Worse than average
Size of gap—Very much larger than average

ASK3L W change—Unavailable
ASK3L B change—Unavailable

ASK3M W change—Better than average
ASK3M B change—Average

ASK4 L W—Average
ASK4 L B—Worse than average
Size of gap—Larger than average

ASK4M W—Average
ASK4M B—Much worse than average
Size of gap—Larger than average

ASK4 S W—Above average
ASK4S B—Much worse than average
Size of gap—Much larger than average

ASK4L W change—Average
ASK4L B change—Much worse than average

ASK4M W change—Very much worse than average
ASK4M B change—Much worse than average

Comments: Seth Boyden is marked by achievement gaps that range from larger to very much larger than the gaps in the DFG. Performance by whites ranges from average to much above average; performance by blacks ranges from worse than average to much worse than average. Change results are spotty for both groups, with both blacks and whites showing a strongly negative change relative to the state on ASK4 M from 2004 to 2005.
g. South Mountain Elementary School

ASK3 L W—Unavailable
ASK3 L B—Unavailable
Size of gap—Unavailable

ASK3M W—Unavailable
ASK3M B—Unavailable
Size of gap—Unavailable

ASK3L W change—Unavailable
ASK3L B change—Unavailable

ASK3M W change—Unavailable
ASK3M B change—Unavailable

ASK4 L W—Average +
ASK4 L B—Above average
Size of gap—Somewhat smaller than average

ASK4M W—Very much better than average
ASK4M B—Much better than average
Size of gap—Somewhat larger than average

ASK4 S W—Much above average
ASK4S B—Above average
Size of gap—Somewhat larger than average

ASK4L W change—Average -
ASK4L B change—Average

ASK4M W change—Above average
ASK4M B change—Very much above average

Comments: South Mountain has achievement gaps relative to the DFG that range from somewhat smaller than average to somewhat larger than average. Performance by whites relative to the DFG ranges from somewhat above average to very much above average; performance by blacks ranges from above average to much above average. Change results are fine for both groups, with blacks showing a strongly positive change relative to the state on ASK4 M from 2004 to 2005. Overall, South Mountain illustrates good performance relative to the DFG by both blacks and whites in a district not characterized by such good performance by both groups relative to the DFG.

h. Tuscan Elementary School

Disaggregated 2005 data unavailable
IV. Practical Judgments and Value Judgments Involved in the Use of Aggregate and Disaggregated Data

After the passage of the No Child Left Behind Act, information has become available for SO-M and other districts on a disaggregated basis that breaks down performance by different subgroups, including racial groups, males and females, and general education and special education students. Disaggregated data are not available, with limited exceptions, to compare subgroups defined by socioeconomic status. Depending on the circumstances, there can be significant practical value to both aggregate data comparisons involving all students and disaggregated data comparisons involving subgroups. There are also significant and complex value judgments attending the use of both kinds of data.

Disaggregated data may usefully focus attention by a community on subgroup performance that especially needs to be improved, and also usefully allows comparisons of school and district performance that are controlled for group composition. But reliance on disaggregation may also contribute to a defeatist attitude toward performance—“we’re going to do only this well as a district given our racial composition”—or a strengthening of negative beliefs about the skills of members of groups in which average performance is weaker. With the passage of NCLB, the U.S. made a decision that disaggregated data had sufficient value to warrant mandating its disclosure. One can believe that that decision was correct and still believe that the handling and discussion of disaggregated data need to be done with great care. It is important that districts not use their racial or other group makeup as an excuse for poor performance, and there needs to be concern about the possibility that disaggregated data will be used in a way that reinforces negative stereotypes. If disaggregation is used as a tool for people to understand school and district performance better and to make positive change, it is to be applauded; if it is used as a crutch for mediocre or poor teaching, administration, or learning, it should be decried.

One analytical point is worth noting: The claim is sometimes made that disaggregation is the only statistically valid way to consider how well schools, as opposed to students’ background characteristics, are affecting performance. That claim is overdrawn, for the following reason: It is correct that controlling for background factors is highly desirable in order to estimate the effect of a given variable of interest, such as teaching, on an outcome, such as test performance. But there is a difficulty, in that controlling for a factor such as race entails an assumption that an effect associated with that factor is not caused by the variable of interest. Race and other background factors such as socioeconomic status are not themselves created by schools, of course—but there is a tricky question as to how much the effects associated with race, SES, and so on are caused by schools as opposed to other factors. Using aggregate data that ignores race or other background factors makes an assumption that the schools are solely responsible for racial or other group effects on performance, while using disaggregated data to control for race or other background factors makes the assumption that the schools are not responsible for racial or other group effects at all. Neither assumption is realistic, which suggests that a
sensible overall approach to data has to include both aggregate data and disaggregated data.

In conclusion: For a school board or other policy-maker, it is inadequate to rely on either aggregate or disaggregated data alone. The richer data universe available after NCLB is a definite net benefit for educators and communities compared to the pre-NCLB data, in which report cards in New Jersey were limited to aggregate data that by itself was often misleading. But post-NCLB it is important not to swing to the opposite extreme of focusing solely on disaggregated data. Aggregate performance and change results are highly important and should not be ignored or rationalized away.
V. Notes on Methods and Assumptions

Note 1: Comparisons of SO-M schools and subgroups in Sections II and III are to the district factor group I average (the group to which SO-M is assigned by the state based on our towns’ socio-economic status), unless it is noted that the particular comparison is to the state as a whole or to another school district or school.

Note 2: In examining the analysis, the reader is asked to bear in mind the complexities of interpreting data and the need for judgment in doing so.

Note 3: The criteria for evaluating test performance as average, average +, average -, above average, below average, much above average, much below average, very much above average, and very much below average relative to DFG performance work as follows:

a. Take the difference between advanced proficient performance and partially proficient performance for the given test;
b. From that difference, subtract the comparable difference between advanced proficient and partially proficient performance for DFG I;
c. If the result of the operations above is from -10 to +10, performance is rated as average, with +4 to +10 as average + and -4 to -10 as average -. If the result is from +10 to +20, performance is rated as above average; from -10 to -20, below average. If the result is from +20 to +30, performance is rated as much above average; from -20 to -30, much below average. If the result is +30 or above, performance is rated as very much above average; if -30 or below, very much below average.

Note 4: The criteria for evaluating change in test performance as average, average +, average -, above average, below average, much above average, much below average, very much above average, and very much below average relative to state performance (DFG data were not available for this purpose) work as follows:

a. Take the difference between advanced proficient performance and partially proficient performance for the given test in 2005 minus the difference between advanced proficient and partially proficient in 2004;
b. From that difference, subtract the comparable difference between advanced proficient and partially proficient performance for the state in 2005 minus the difference between advanced proficient and partially proficient in 2004;
c. If the result of the operations above is from -5 to +5, performance on the change factor is rated as average, with +3 to +5 as average + and -3 to -5 as average -. If the result is from +5 to +15, performance is rated as above average; from -5 to -15, below average. If the result is from +15 to +25, performance is rated as much above average; from -15 to -25, much below average. If the result is +25 or above, performance is rated as very much above average; if -25 or below, very much below average.
Note 5: SRA versus HSPA graduation comparisons are made in absolute percentage points (so, for example, Columbia’s 18% SRA rate compared to the state’s 14% rate is a 4 point difference.)

Note 6: SAT comparisons are made in points, using ETS’s 200-800 scale. 2005 SAT comparisons are to the DFG; SAT change comparisons are to the state.

Note 7: AP participation comparisons with the state are made in absolute percentage points.

Note 8: The ratio of the number of AP scores of 3 or above to the school population was computed by dividing the number of 3+ results by the total number of tests taken and multiplying that by the percentage of Grade 11-12 students participating.

Note 9: Comparisons of SAT participation between high schools are made in absolute percentage points.

Note 10: Comparisons of AP scores of 3 or over between high schools are made in absolute percentage points.

Note 11: In comparing performance in Abbott district schools to performance in SO-M schools, the following criteria were used:
  a. If the difference between two schools’ advanced minus partially proficient % is 2 points or greater, the school with the better difference is considered better on that test;
  b. If the difference between two schools’ advanced minus partially proficient % is from 0-2 points, the schools are considered comparable on that test.

Note 12: The point comparisons for the HSPA between Columbia and Livingston, Millburn, Montclair, and West Orange are done as follows:
  a. Take the difference between advanced proficient performance and partially proficient performance for Columbia on the given test;
  b. From that difference, subtract the comparable difference between advanced proficient and partially proficient performance for the other school on the given test.

Note 13: The achievement gap between blacks and whites is defined in terms of size relative to the gap in the DFG—average (0-5 point deviation from DFG), somewhat smaller (or larger) than average (5-10 point deviation from DFG), smaller (or larger) than average (10-20 point deviation from DFG), much smaller (or larger) than average (20-30 point deviation from DFG), and very much smaller (or larger) than average (30 point or more deviation from DFG).